Growth Charts – Curvas de Crecimiento

Las curvas de crecimiento infantil son herramientas fundamentales para la evaluación clínica y epidemiológica del desarrollo físico. Sin embargo, las curvas más utilizadas en México, como las del CDC 2000, no han sido actualizadas en más de dos décadas y fueron construidas con datos de población estadounidense que no reflejan las condiciones nutricionales, genéticas y sociales de la infancia mexicana actual. Esta obsolescencia puede llevar a errores diagnósticos, especialmente durante la pubertad, etapa en la que la variabilidad en la maduración biológica no es capturada por modelos basados exclusivamente en la edad cronológica. Por ello, esta tesis propone la construcción de nuevas curvas de crecimiento ajustadas al contexto mexicano, utilizando datos de la Encuesta Nacional de Salud y Nutrición (ENSANUT) y modelado estadístico bayesiano. Se busca integrar la edad, el sexo, y —en la medida de lo posible— indicadores de maduración puberal y hábitos de vida, con el fin de obtener una herramienta más precisa y contextualizada para la evaluación del crecimiento infantil en México.

Las curvas de crecimiento infantil del CDC (Centers for Disease Control and Prevention) publicadas en el año 2000 son una de las principales herramientas utilizadas a nivel internacional para evaluar el crecimiento físico de niños y adolescentes de 0 a 20 años. Estas curvas se construyeron a partir de datos recopilados entre 1963 y 1994 a través de cinco encuestas nacionales en Estados Unidos (NHES II-III, NHANES I-III y HHANES). Incluyeron información de niños blancos no hispanos, negros no hispanos y una submuestra de niños mexicano-americanos nacidos y criados en el suroeste de Estados Unidos.

El método estadístico utilizado fue el modelo LMS (Lambda-Mu-Sigma), que permite modelar la distribución de variables como la talla, el peso y el índice de masa corporal (IMC) en función de la edad, mediante funciones suavizadas de la mediana, la variabilidad y la asimetría. Las curvas resultantes se expresan típicamente en forma de percentiles (P3, P15, P50, P85, P97) o puntuaciones z.

Aunque estas curvas han sido adoptadas ampliamente, presentan limitaciones importantes para su aplicación en poblaciones como la mexicana. En primer lugar, no se basan en datos de niños residentes en México, sino en una población específica de mexicano-americanos que difiere nutricional y socioeconómicamente. Además, las curvas del CDC no han sido actualizadas en más de dos décadas, a pesar de los cambios sustanciales en los patrones de crecimiento infantil asociados a la transición nutricional, el entorno social y la carga de enfermedades crónicas. Asimismo, no incorporan variables clave como el estadio puberal o los hábitos de vida, lo que limita su precisión diagnóstica durante la adolescencia.

Estas limitaciones justifican la necesidad de desarrollar curvas de crecimiento ajustadas al contexto actual de México, que integren factores biológicos y conductuales, y que utilicen modelos estadísticos contemporáneos para una evaluación más adecuada del crecimiento infantil.

<https://www.cdc.gov/growthcharts/cdc-charts.htm>

**📘 Estructura – con ayuda de Chat GPT**

**1. Título tentativo**

Estimación de la maduración puberal y modelado bayesiano del crecimiento infantil en México utilizando datos de ENSANUT y curvas de referencia del CDC

**2. Planteamiento del problema**

* Las curvas tradicionales de crecimiento (CDC/OMS) no consideran el estadio puberal, lo que lleva a posibles errores diagnósticos.
* México no cuenta con curvas ajustadas a su realidad epidemiológica y puberal.
* La ENSANUT no incluye directamente el estadio Tanner, pero sí variables correlacionadas.
* Objetivo: predecir el estadio puberal con ML y usarlo como covariable latente en un modelo bayesiano del crecimiento.

**3. Hipótesis**

* H1: Es posible construir un proxy confiable del estadio puberal usando ML sobre variables ENSANUT.
* H2: Incluir ese proxy en un modelo bayesiano mejora la predicción y explicación de la estatura respecto al uso exclusivo de edad cronológica.

**4. Objetivos**

**Objetivo general:**

Construir un modelo bayesiano del crecimiento infantil ajustado por estadio puberal estimado mediante aprendizaje automático.

**Objetivos específicos:**

1. Explorar las variables ENSANUT asociadas a maduración puberal.
2. Construir un modelo de clasificación supervisado para predecir estadios puberales.
3. Validar el proxy con literatura externa sobre edad esperada de Tanner.
4. Integrar el proxy en un modelo jerárquico bayesiano (p. ej. log-normal o LMS).
5. Comparar resultados con curvas CDC y analizar sesgos de clasificación en la población mexicana.

**5. Marco teórico (resumido)**

**Crecimiento infantil**

* Fisiología del crecimiento.
* Pubertad y eje HPG.
* Etapas Tanner.

**Curvas de crecimiento**

* CDC 2000, OMS 2007.
* Limitaciones para México.
* Curvas ajustadas por pubertad (TSAHeight).

**Machine learning para variables latentes**

* Random Forest / XGBoost como estimadores de proxies.
* Validación cruzada, feature importance.

**Modelos bayesianos**

* Modelado de crecimiento con LMS, GAMLSS o regresión log-normal jerárquica.
* Uso de PyMC, Stan o brms para inferencia bayesiana.
* Priori informativa vs no informativa.

**6. Datos**

**Fuente:**

* ENSANUT (probablemente 2012 o 2021 dependiendo del acceso).
* Muestra: niños y adolescentes de 8–18 años.
* Variables clave:
  + Edad, sexo, estatura, peso, IMC
  + Actividad física, alimentación, estado nutricional

**7. Metodología**

**A. Preprocesamiento**

* Limpieza y normalización
* Z-scores CDC
* Transformaciones necesarias para el modelo bayesiano

**B. Estimación del proxy puberal**

1. Selección de variables correlacionadas
2. Algoritmos: Random Forest, Gradient Boosting o LCA
3. Validación: AUC, matriz de confusión, plausibilidad biológica (edad esperada por sexo)

**C. Modelado bayesiano**

1. Variable dependiente: estatura (talla en cm)
2. Covariables:
   * Edad
   * Sexo
   * Proxy de estadio puberal (como ordinal o categórica)
3. Modelo propuesto: regresión jerárquica log-normal o LMS bayesiano
4. Software: PyMC, Stan o brms (en R)
5. MCMC, trazas, verificación de convergencia
6. Intervalos de credibilidad para proyecciones de talla

**8. Resultados esperados**

* Mapa nacional de distribución estimada de estadio puberal.
* Comparación de predicciones de estatura con y sin proxy.
* Visualización de curvas ajustadas (tipo TSAHeight para México).
* Discusión sobre política pública: sesgo de talla baja o pubertad temprana.

**9. Conclusiones**

* Valor agregado de los métodos ML y bayesianos en epidemiología infantil.
* Posibilidades futuras para curvas de crecimiento mexicanas.
* Necesidad de incluir evaluación clínica del estadio Tanner en ENSANUT o encuestas nacionales.